§2.2 Organizing Data (frequency distributions)

Example of Categorical Frequency Distribution

A sample of 20 M&M's is observed and their colors are recorded.

Green	Yellow
Red	Green
Yellow	Brown
Red	Brown
Brown	Yellow
Blue	Red
Red	Orange
Green	Yellow
Red	Red
Yellow	Green

M&M Color	Sample Frequency	Percent
Yellow	5	25%
Green	4	20%
Red	6	30%
Brown	3	15%
Orange	1	5%
Blue	1	5%
Total	20	100%

Categorical frequency distribution.

Group Frequency Distribution

Defining the classes (or groups):

- **Class limits,** for a given class, are the lowest and highest data values that are included in a class (should have same number of decimal places as recorded data).
- **Class boundaries** are defined to eliminate any gaps between the classes (has one more decimal place that data.)
- **Class width** is lower limit of class subtracted from lower limit of next class

Times for 400 meter race

45.38	45.41
51.52	48.79
49.72	47.28
47.03	47.16
47.52	51.60
50.66	45.25
47.23	46.95
47.70	46.34
51.15	50.74
47.63	48.23
48.72	51.57
49.44	47.24
46.43	

Class Limits	Class Boundaries	Frequency	Cumulative Freq.
45.25 - 46.30	45.245 - 46.305	3	3
46.31 - 47.36	46.306 - 47.365	8	11
47.37 - 48.42	47.365 - 48.425	4	15
48.43 - 49.48	48.425 - 49.485	3	18
49.49 - 50.54	49.485 - 50.545	1	19
50.55 - 51.60	50.545 - 51.605	6	25

Group Frequency Distribution

3

Rules (properties) for setting up classes for group freq. dist.

1. Should be from 5 to 20 classes (rule of thumb).
2. Class width should be an odd number (suggestion).
3. Classes must be mutually exclusive.
4. Classes must be continuous.
5. Classes must be exhaustive.
6. Classes must have equal width. (Except for an open ended class.)

Times for 400 meter race

45.38	45.41
51.52	48.79
49.72	47.28
47.03	47.16
47.52	51.60
50.66	45.25
47.23	46.95
47.70	46.34
51.15	50.74
47.63	48.23
48.72	51.57
49.44	47.24
46.43	

Class Limits	Class Boundaries	Frequency	Cumulative Freq.
45.25 - 46.30	45.245 - 46.305	3	3
46.31 - 47.36	46.306 - 47.365	8	11
47.37 - 48.42	47.365 - 48.425	4	15
48.43 - 49.48	48.425 - 49.485	3	18
49.49 - 50.54	49.485 - 50.545	1	19
50.55 - 51.60	50.545 - 51.605	6	25

4

Steps for setting up the classes (see pages 38 & 40)

1. Determine the classes
   a) Find the range: Range = Highest - Lowest
   b) Select the number of classes (5 to 20)
   c) Calculate width by dividing range by number of classes and rounding **up** (to same # of significant digits as recorded data).
   d) Select a starting point for the lowest class limit and calculate all lower class limits.
   e) Find the upper class limits.
   f) Find the boundaries.
2. Calculate numerical frequencies.
3. Calculate cumulative frequencies.
4. Calculate midpoints of classes.

5

Example of Group Frequency Distribution

Blood glucose levels for sample of 60 patients

55	115	118	114	59	109
63	97	90	59	105	81
84	81	82	61	103	77
82	76	68	86	97	80
77	85	69	62	101	83
58	83	101	86	84	78
59	92	88	97	87	92
70	86	72	84	82	84
101	80	93	56	65	91
75	78	100	74	74	90

Create group frequency distribution with **6** groups.

6

Example of Group Frequency Distribution

Blood glucose levels for sample of 60 patients

55	115	118	114	59	109
63	97	90	59	105	81
84	81	82	61	103	77
82	76	68	86	97	80
77	85	69	62	101	83
58	83	101	86	84	78
59	92	88	97	87	92
70	86	72	84	82	84
101	80	93	56	65	91
75	78	100	74	74	90

Class Limits	Class Boundaries	Frequency	Percent	C. Freq.	C. %
55 - 65	54.5 - 65.5	10	16.7%	10	16.7%
66 - 76	65.5 - 76.5	8	13.3%	18	30.0%
77 - 87	76.5 - 87.5	22	36.7%	40	66.7%
88 - 98	87.5 - 98.5	10	16.7%	50	83.3%
99 - 109	98.5 - 109.5	7	11.7%	57	95.0%
110 - 120	109.5 - 120.5	3	5.0%	60	100.0%

7

§2.3 Histograms, Frequency Polygons, and Ogives

The **histogram** is a graph that displays the data by using contiguous vertical bars of various heights to represent the frequencies of the classes. (class boundaries along *x* axis)

The **frequency polygon** is a graph that displays the data by using lines that connect points plotted for the frequencies, at the midpoints of the classes. (class midpoints along *x* axis)

The **ogive** (cumulative frequency graph) is a graph that shows the cumulative frequencies for the classes. (with connected points and class boundaries along x axis).

Class Limits	Class Boundaries	Frequency	Percent	C. Freq.	C. %
55 - 65	54.5 - 65.5	10	16.7%	10	16.7%
66 - 76	65.5 - 76.5	8	13.3%	18	30.0%
77 - 87	76.5 - 87.5	22	36.7%	40	66.7%
88 - 98	87.5 - 98.5	10	16.7%	50	83.3%
99 - 109	98.5 - 109.5	7	11.7%	57	95.0%
110 - 120	109.5 - 120.5	3	5.0%	60	100.0%

8

§2.3 Relative Frequency Graphs

- **Relative Frequency Histogram**
- **Relative Frequency Polygon**
- **Cumulative Relative Frequency ("ogive") Graph**

Class Limits	Class Boundaries	Frequency	Percent	C. Freq.	C. %
55 - 65	54.5 - 65.5	10	16.7%	10	16.7%
66 - 76	65.5 - 76.5	8	13.3%	18	30.0%
77 - 87	76.5 - 87.5	22	36.7%	40	66.7%
88 - 98	87.5 - 98.5	10	16.7%	50	83.3%
99 - 109	98.5 - 109.5	7	11.7%	57	95.0%
110 - 120	109.5 - 120.5	3	5.0%	60	100.0%

Use these instead

9